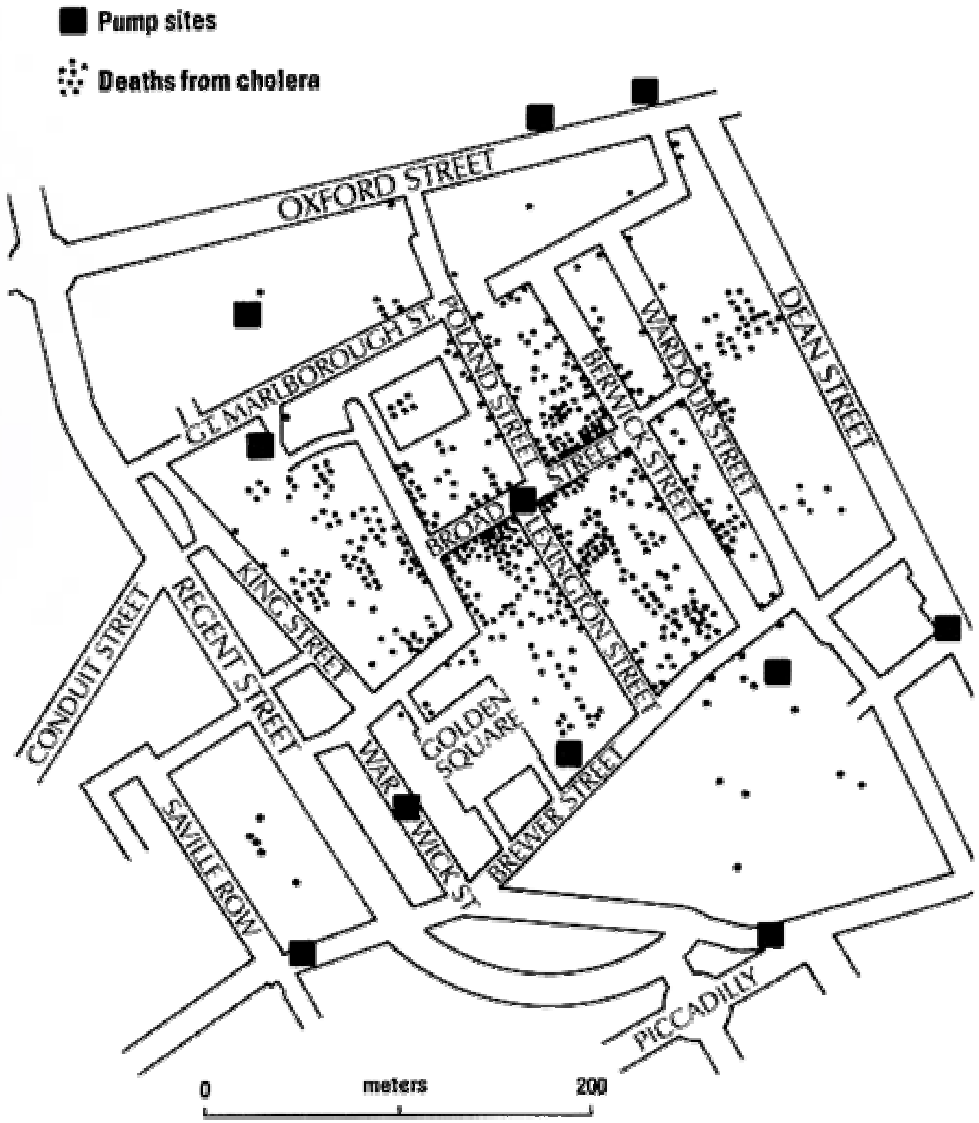
Geographic Information Systems (GIS) as a Tool for Exposure Assessment and Disease Mapping

Verónica Vieira, D.Sc.
Boston University School of Public Health
Department of Environmental Health

John Snow and Cholera Maps



Overview of GIS

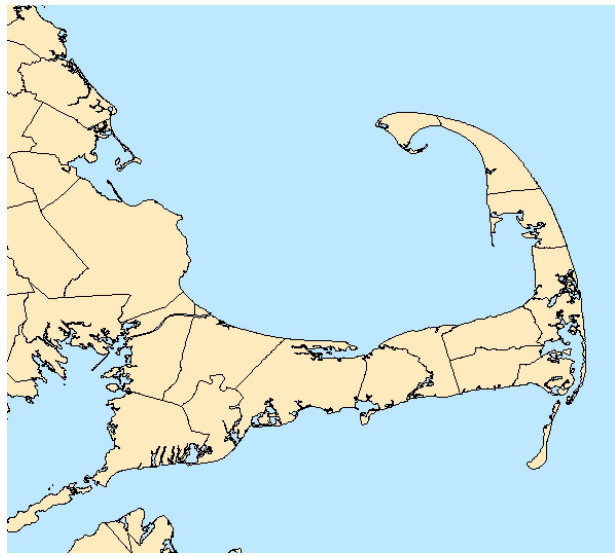
- GIS is a powerful computer mapping and analysis tool that allows environmental and epidemiologic data to be stored, viewed, and analyzed within a geographic context.

- In a GIS map, geographic features are represented by different data layers.

GIS maps contain data layers

- **Vector Data:**

  geographic features represented by shapes (points, lines, or polygons)



- **Raster Data:**

  surface of equally sized cells of square area containing measured or estimated values at that location (elevation, aerial photographs)

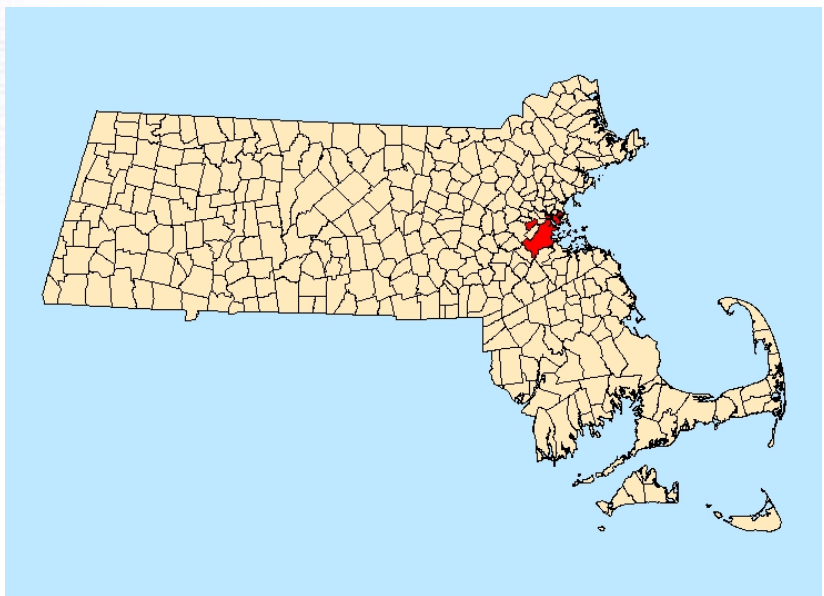
Layers have a coordinate system

- A coordinate system defines locations on the curved surface of the earth.
- X, Y Coordinates are based on the transformation from latitude and longitude.
- GIS can project maps in different coordinates systems.

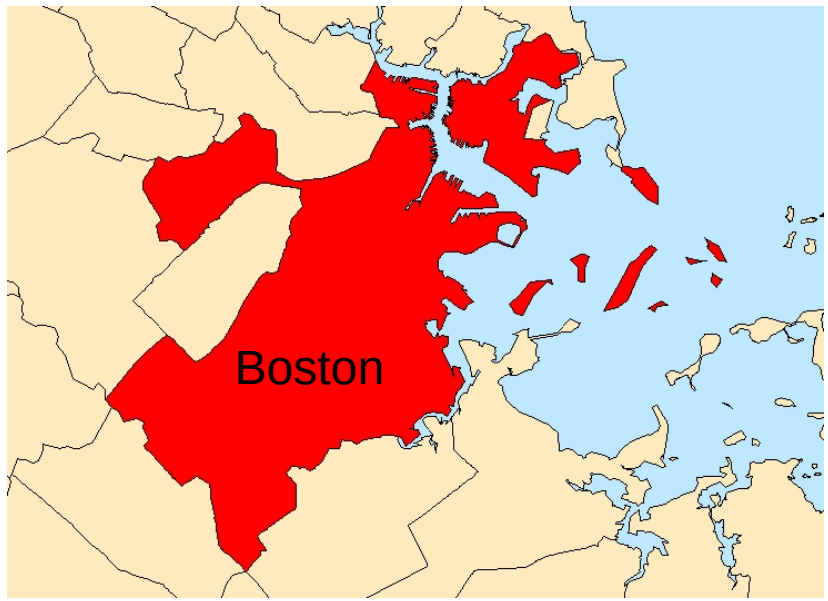
Layers can be different sizes

- Scale is the ratio between the size of features on a map and the size of the features in reality.

1:1,677,543



1:148,577



Layers are linked to data

- Attributes (tabular data) are included with a vector data layer.
- Each layer may contain multiple features.
  - In a single data layer of Massachusetts, each town is a different feature.
- In the attribute table, each feature is a different *record* (row).
- Each record contains a number of *fields* (columns) that store attribute data.

Data stored in attribute tables

- Attributes of the Massachusetts layer include Population in 1980 and 1990.
- Can add and delete fields, sort, select, calculate, join tables.

Attributes of Massachusetts

Shape*	OBJECTID	TOWN_ID	TOWN	POP1980	POP1990	PO
Polygon	1	1	ABINGTON	13579	13817	
Polygon	2	2	ACTON	17672	17872	
Polygon	3	3	ACUSHNET	8808	9554	
Polygon	4	4	ADAMS	10181	9445	
Polygon	5	5	AGAWAM	26754	27323	
Polygon	6	6	ALFORD	402	418	
Polygon	7	7	AMESBURY	14563	14997	
Polygon	8	8	AMHERST	32804	35228	
Polygon	9	9	ANDOVER	27203	29151	

Record: 1 Show: All Selected Records (1 out of 351 Selected.)

GIS can integrate databases

- Overlay several map layers to better understand spatial relationships among the features of different layers.
- Spatially join databases in which the common identifier is spatial location.
- Select features that satisfy a spatial query.

Basic geographic processes

- Dissolve features based on a common attribute
- Merge layers together
- Clip one layer based on another
- Calculate the Intersect of two layers
- Calculate the Union of two layers

GIS & Exposure Assessment

GIS has various applications in exposure assessment, both at an ecologic and individual level.

- Create a buffer around a point source
- Model groundwater plumes
- Estimate environmental levels of contaminants based on measured values
- Estimate personal exposure

Tools for Exposure Assessment

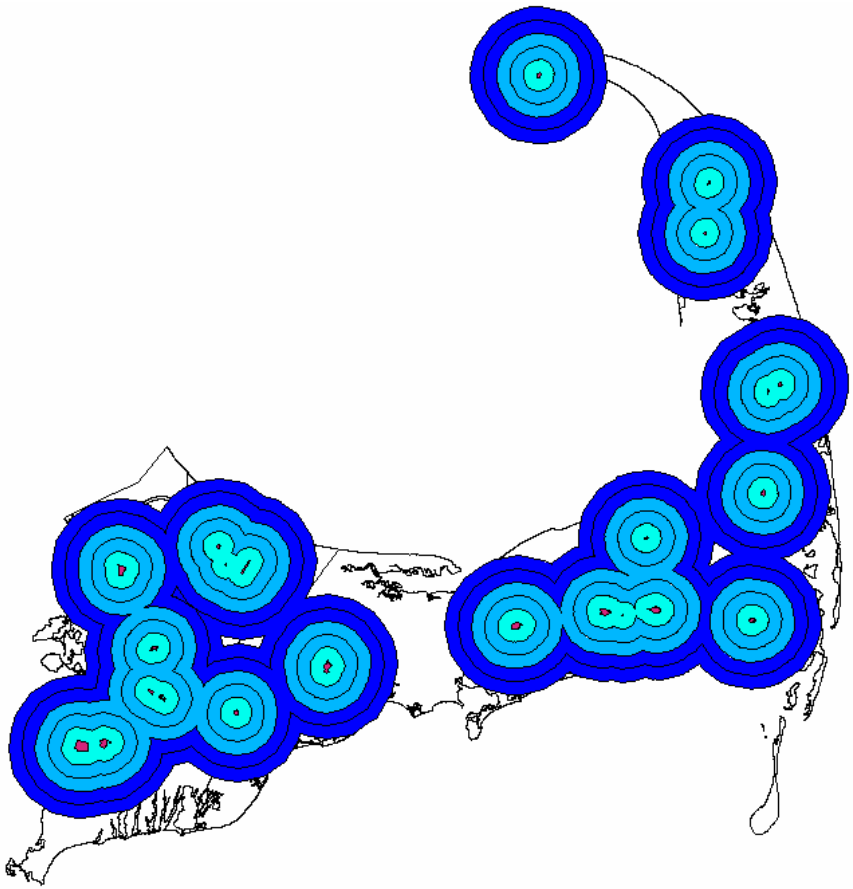
- Automated Address Geocoding
- Buffer Analysis
- Interpolation
- Groundwater Model
- Model Add-Ons

Automated Address Geocoding

- GIS matches the street name in the address to the same street name in the reference (TIGER/Line) files and interpolates within the number range.
- Match rates tend to be low in rural areas

Buffer Analysis

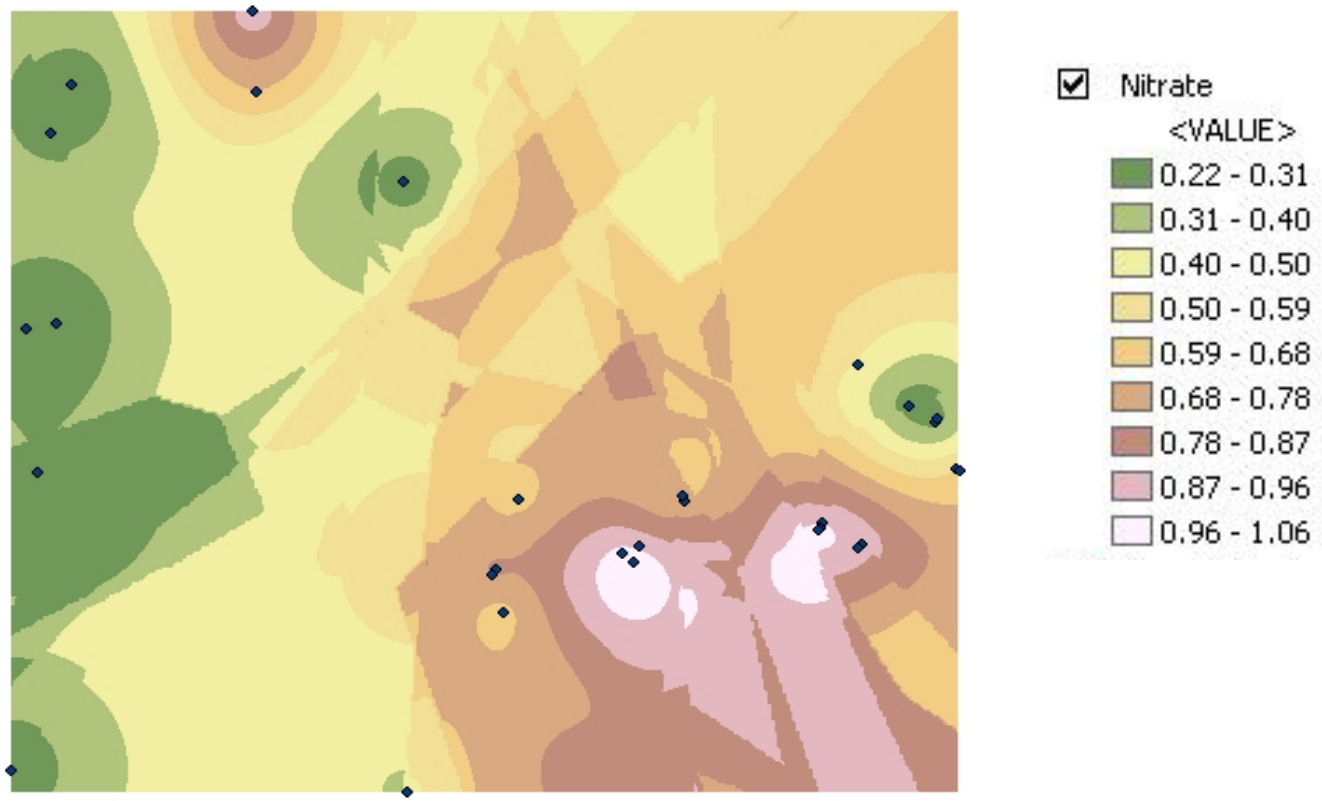
- A buffer of 5 km is drawn around waste disposal sites in intervals of 1 km.



Data Source: MassGIS

Interpolation

- Kriging interpolates between measured nitrate values at public wells to predict a continuous surface of values.



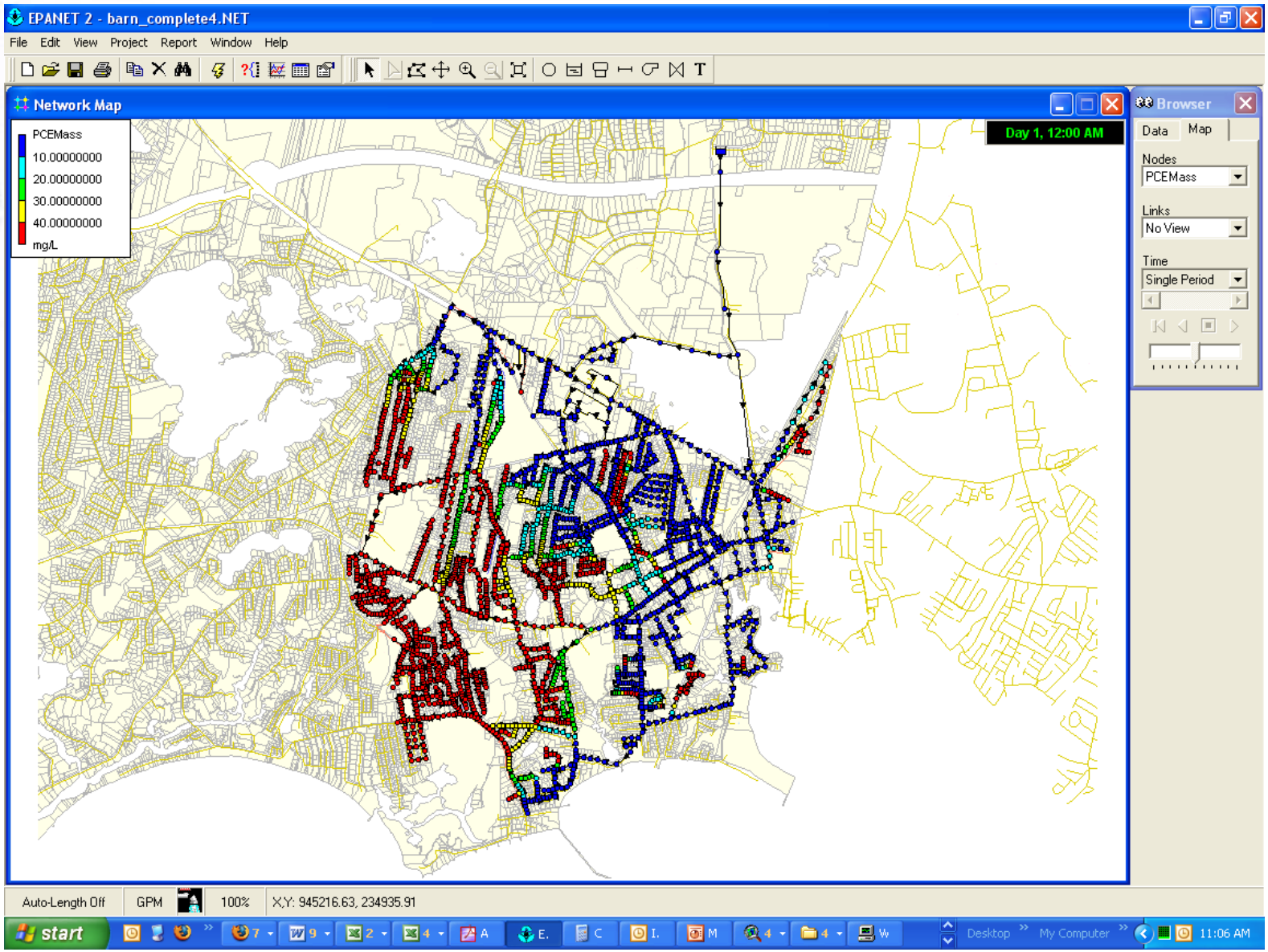
Data Source: MassGIS

Groundwater Tool

- The Particle Track & Porous Puff tools model groundwater plumes from waste disposal sites.

Data Source: MassGIS

Model Add-Ons: EPANET



Considerations

- Data accuracy
- Knowledge of environmental science
- Limitations of measured values
- Model validation
- Geophysical plausibility
- Confounding

Modifiable Areal Unit Problem

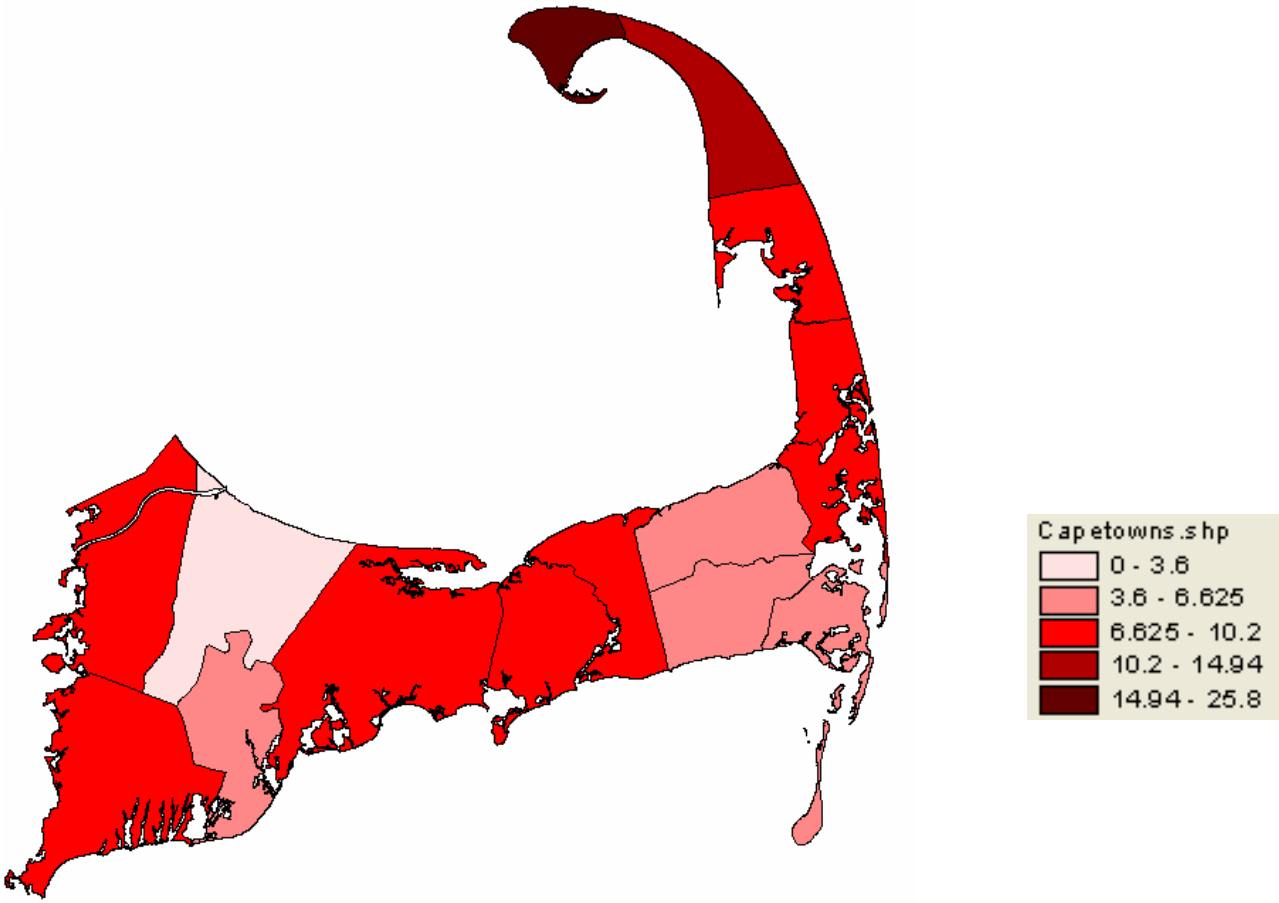
- The level at which you aggregate data changes your results.
- Also affects the ability to detect a relationship between environmental exposure and health outcome.

Levels of Geographic Data

- City/Town
- County
- Zip Code
- Census Tract
- Census Block
- Census Block Groups
- Enumeration Districts
- Individual

Town Level

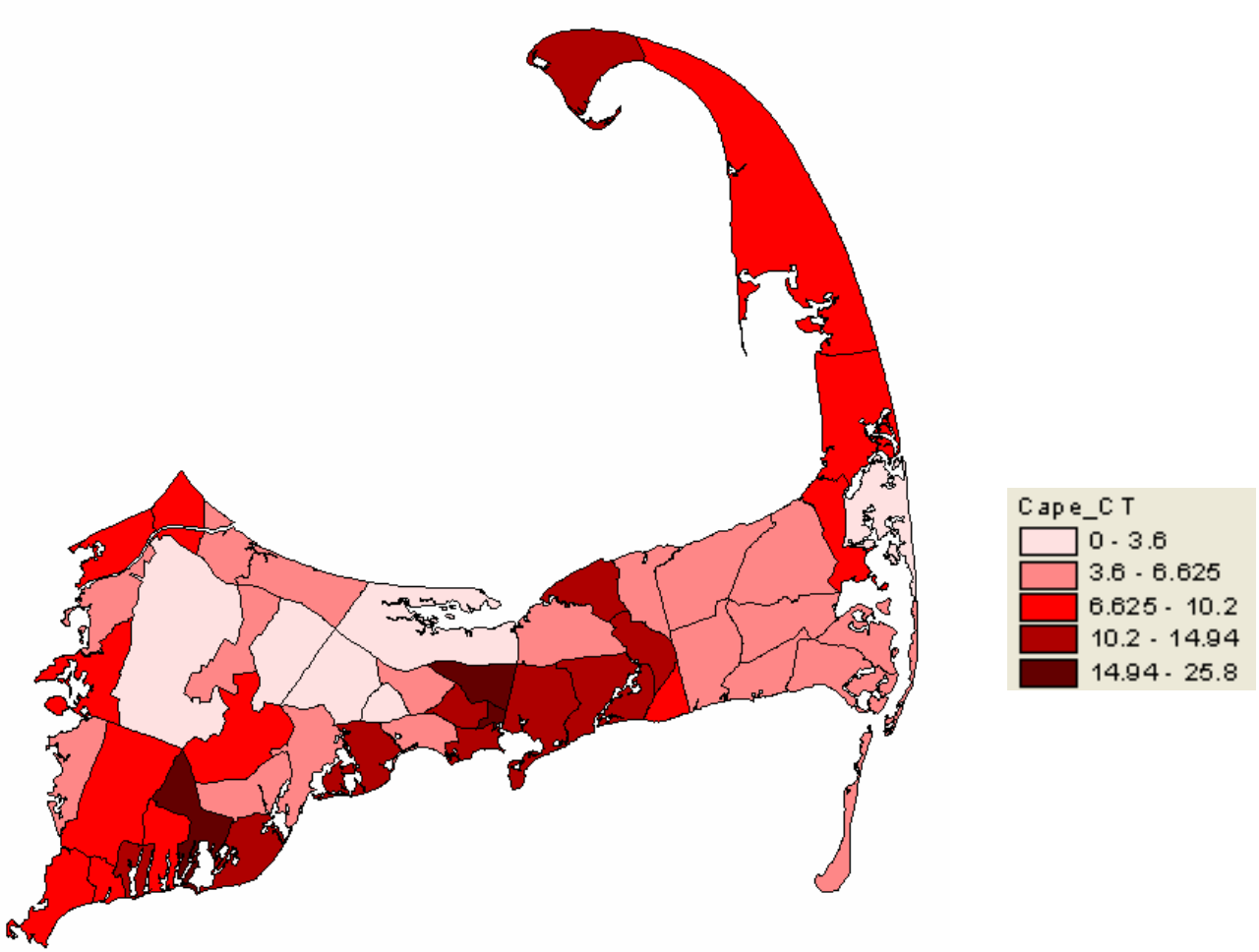
Percent Population below Poverty Level, 1990



Data Source: MassGIS, U.S. Census 1990

Census Tract Level

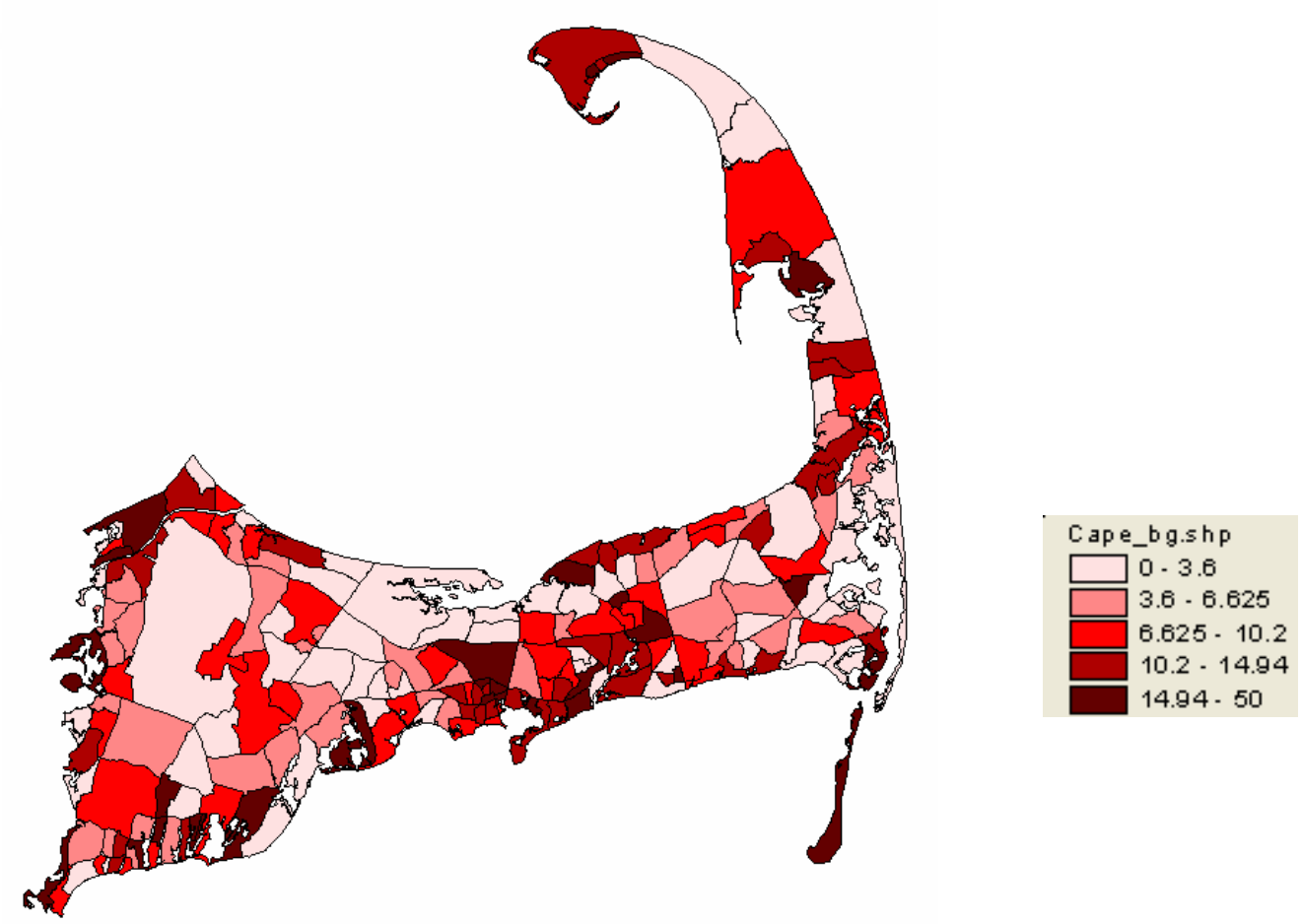
Percent Population below Poverty Level, 1990



Data Source: MassGIS, U.S. Census 1990

Census Block Group Level

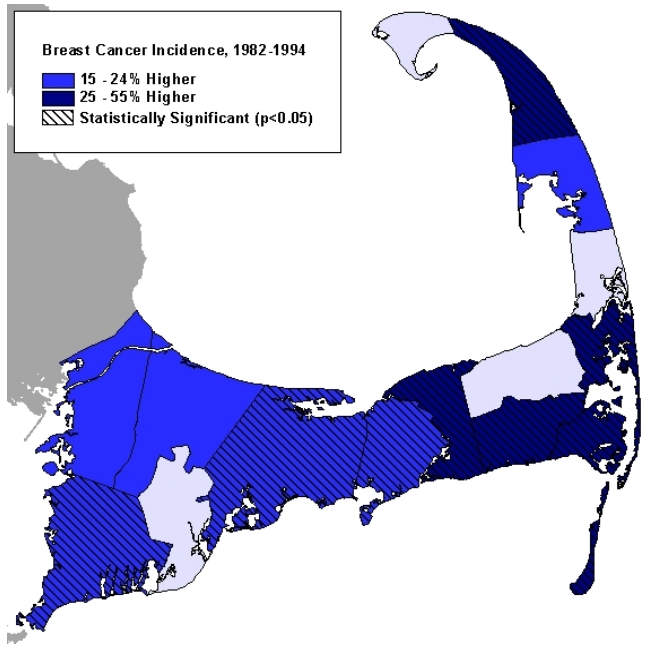
Percent Population below Poverty Level, 1990



Data Source: MassGIS, U.S. Census 1990

GIS & Disease Mapping

- Cancer registry maps
  - arbitrary boundaries
  - ignore latency and possible spatial confounding

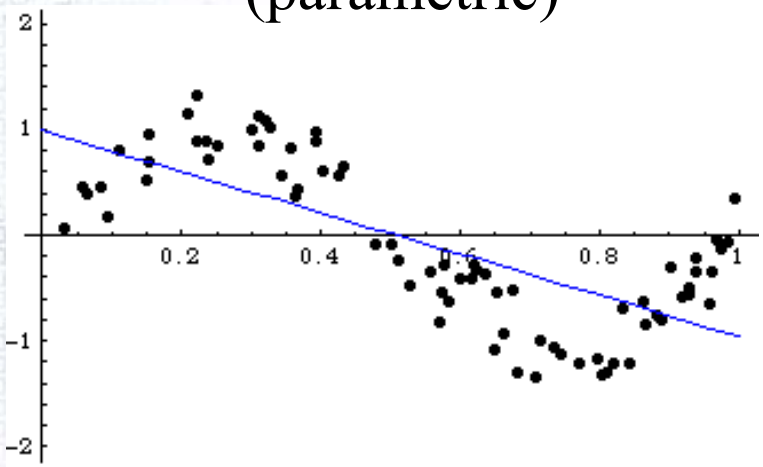


Data Source: Silent Spring Institute

- Applying non-parametric methods to population-based case-control data is one method for dealing with these issues

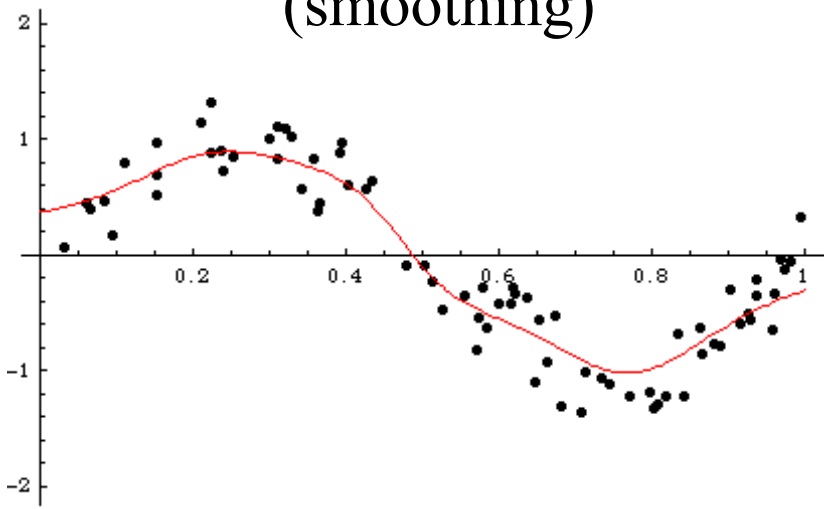
Non-parametric Regression

ordinary linear regression
(parametric)



$y = \alpha + \beta x + \varepsilon$

nonparametric regression
(smoothing)

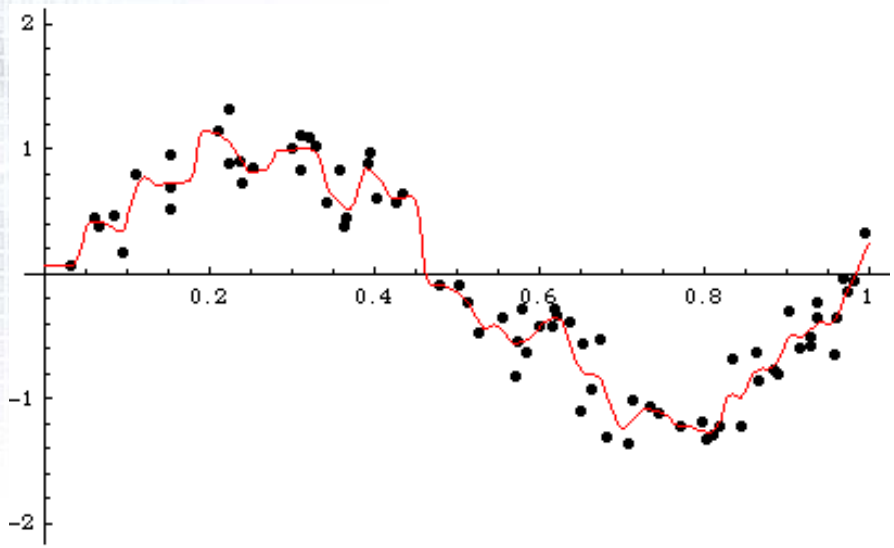


$y = S(x) + \varepsilon$

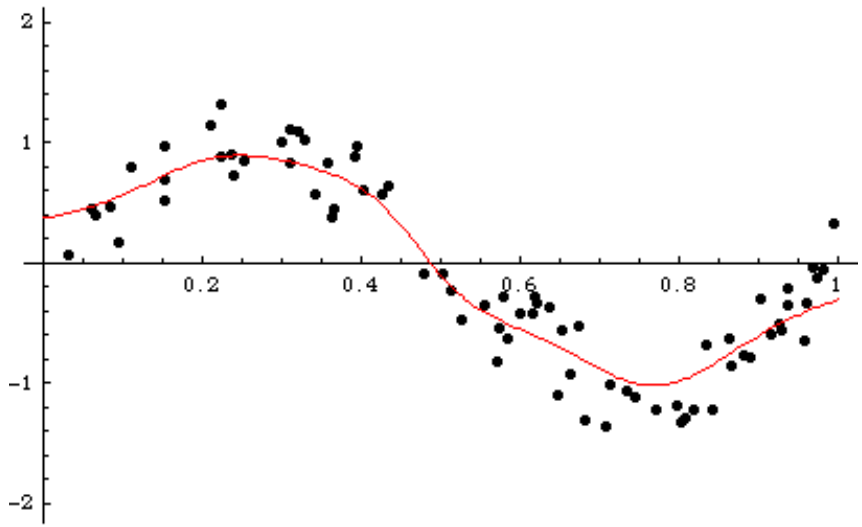
e.g., weighted averages in
"small" neighborhoods

Smoothing

Small smoothing window



Large smoothing window



Generalized Additive Models

$$\text{logit}[p(x1,x2)] = S(x1,x2) + \alpha$$

log disease odds at location (x_1,x_2)

bivariate smoothing function of location

adjusts for covariates where $\boldsymbol{\alpha} = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_j z_j$

GAMs allow for hypothesis testing

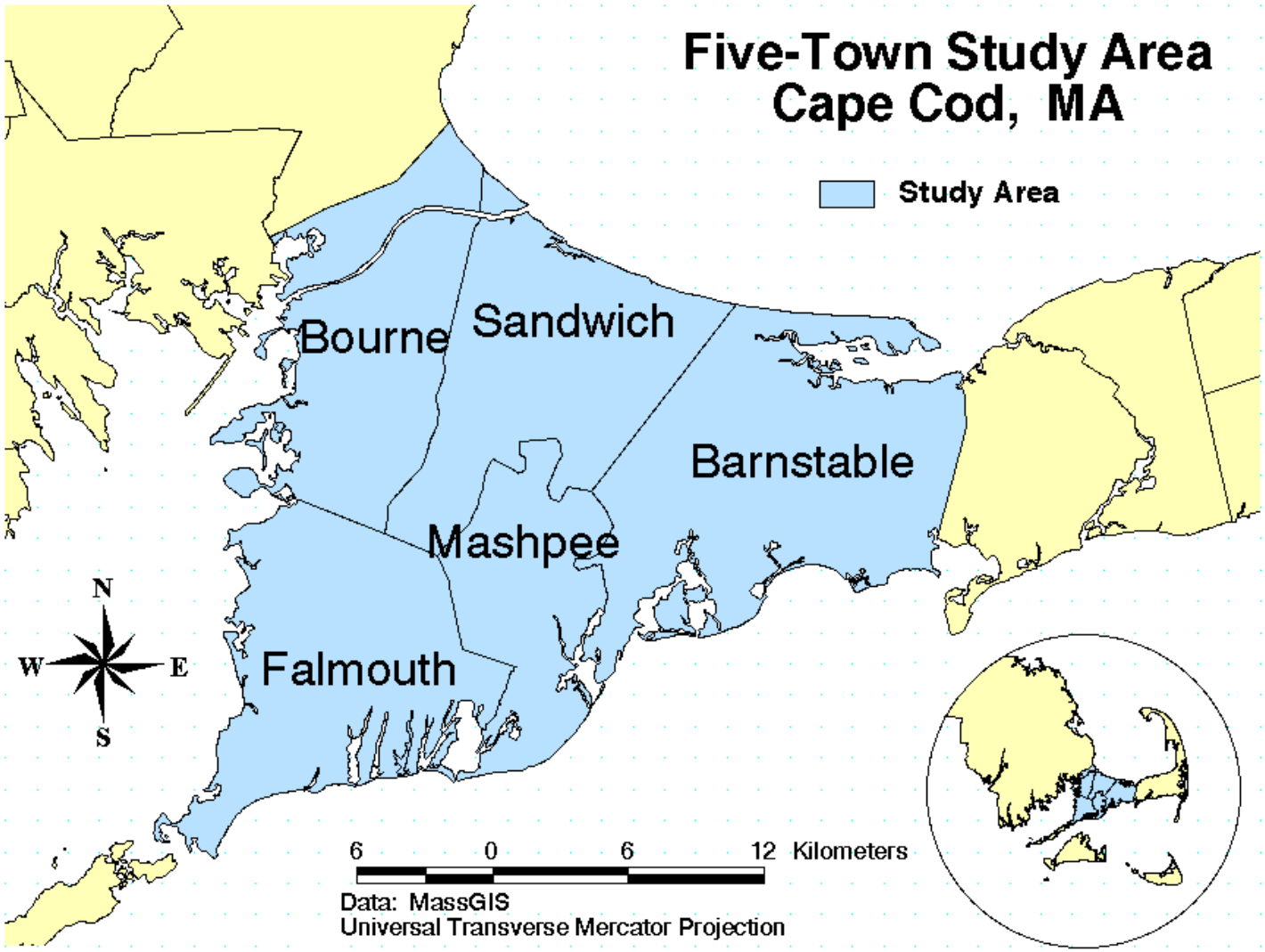
Background

Cape Cod Times, Hyannis, MA. Jan. 5, 1997

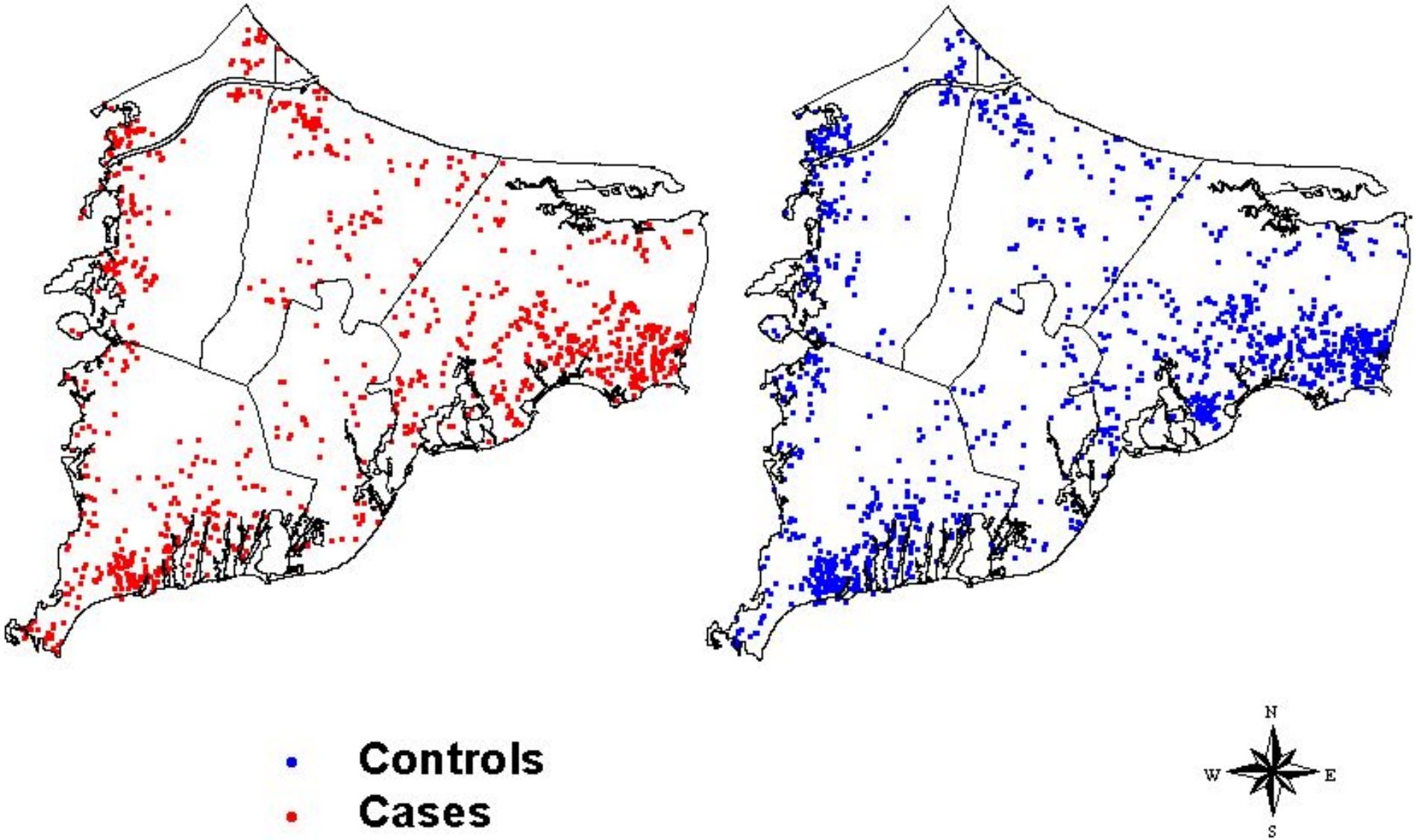
Study Population

- Upper Cape Cancer Study (1989)
- Women's Health on Cape Cod Study (1997)
- Permanent residents, 1983-1993
- Extensive questionnaire
- Vieira et al. (2005)

Study Area


Five-Town Study Area
Cape Cod, MA
Study Area
Bourne
Sandwich
Barnstable
Mashpee
Falmouth
N
W
E
S
6 0 6 12 Kilometers
Data: MassGIS
Universal Transverse Mercator Projection

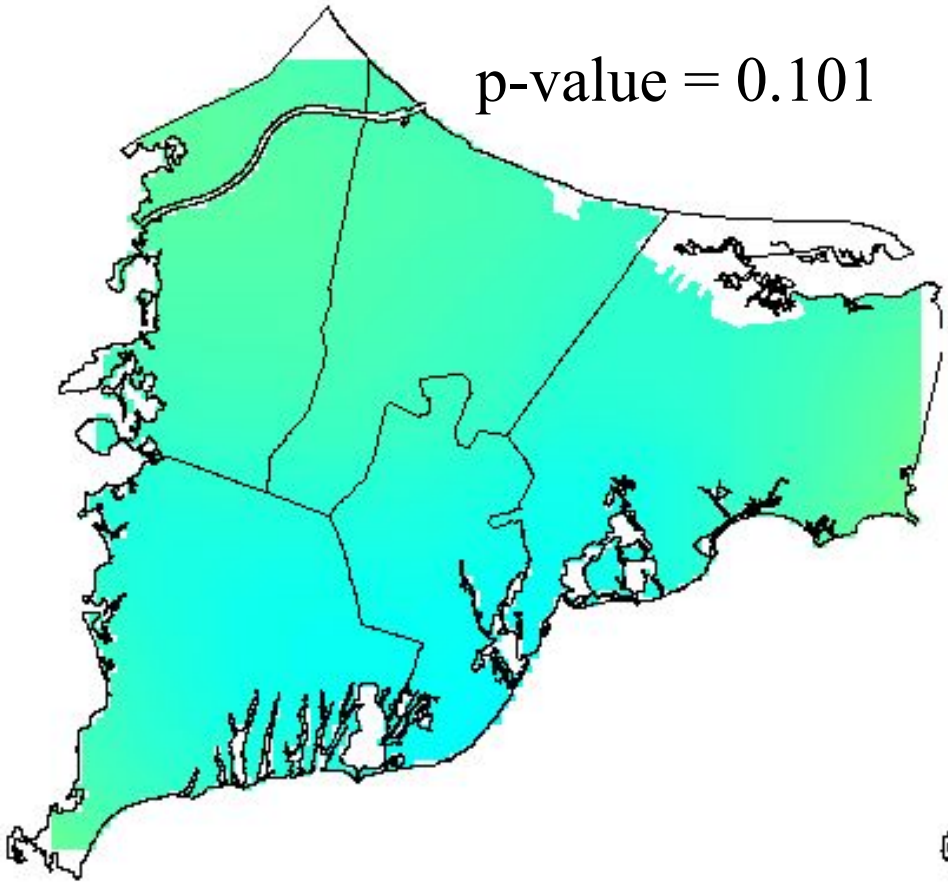
Distribution of Breast Cancer Subjects



Mapping Methods

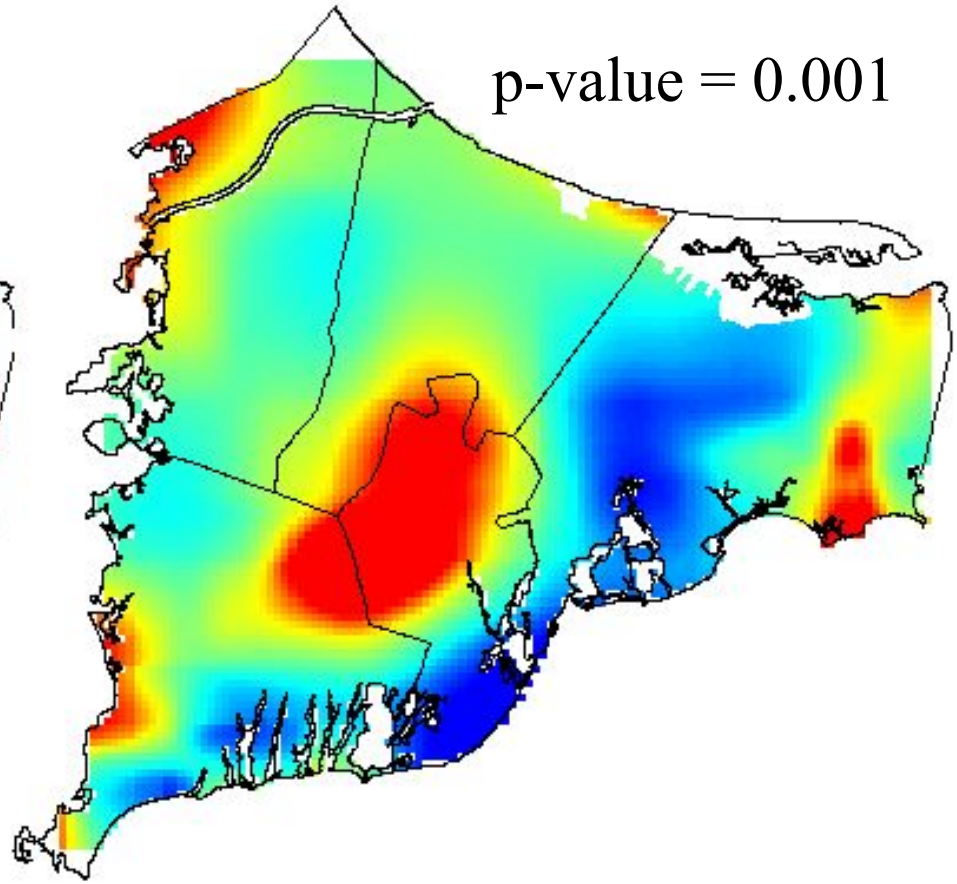
- Odds ratios are relative to the study area
- Blue-to-red color scale range from odds ratios of 0.25 to 2.50
- Model not predicted for wildlife refuge, ocean

Effect of Latency



p-value = 0.101

No Latency

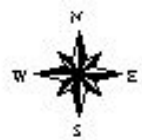


p-value = 0.001

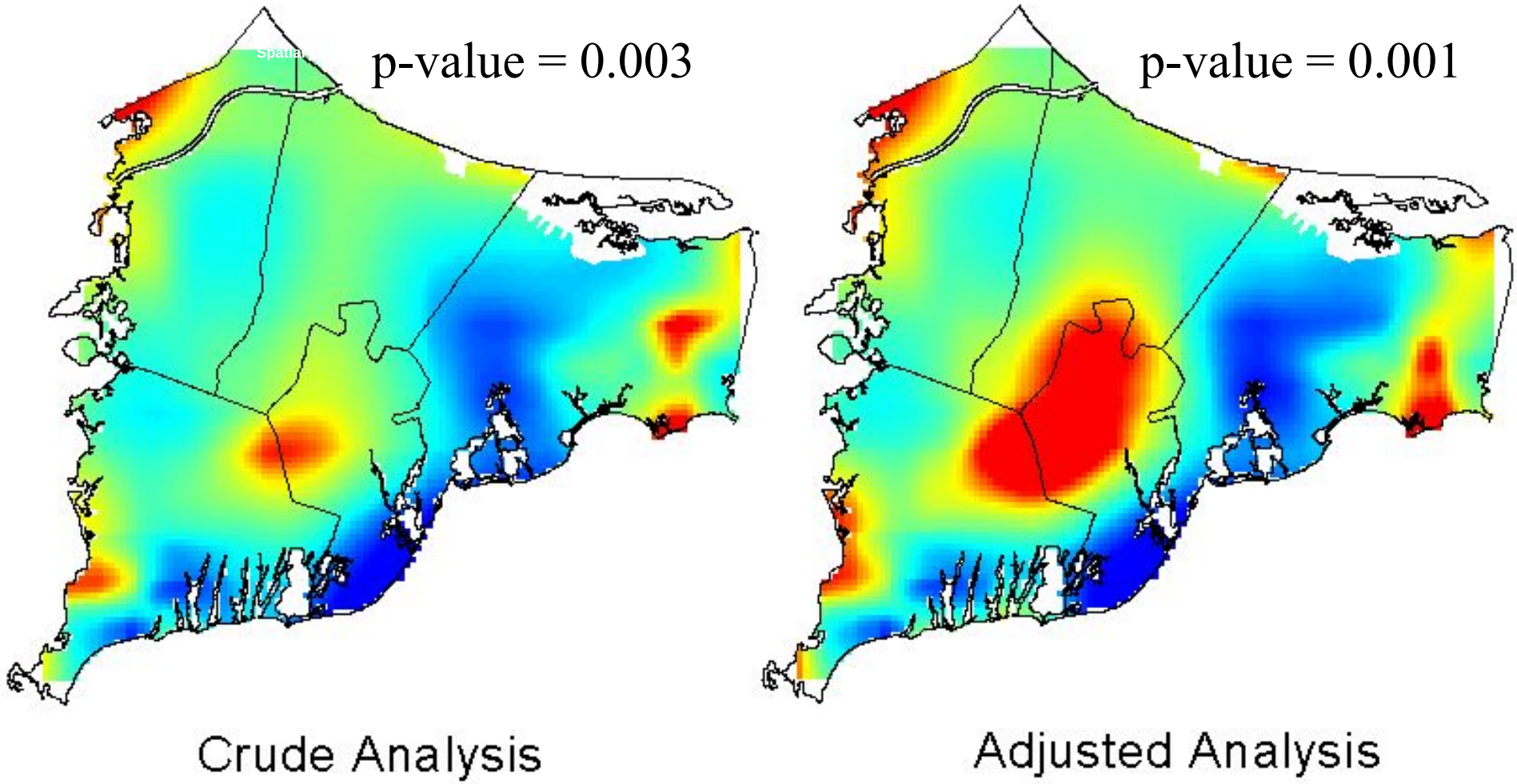
20 Year Latency

Adjusted
Odds Ratio Scale

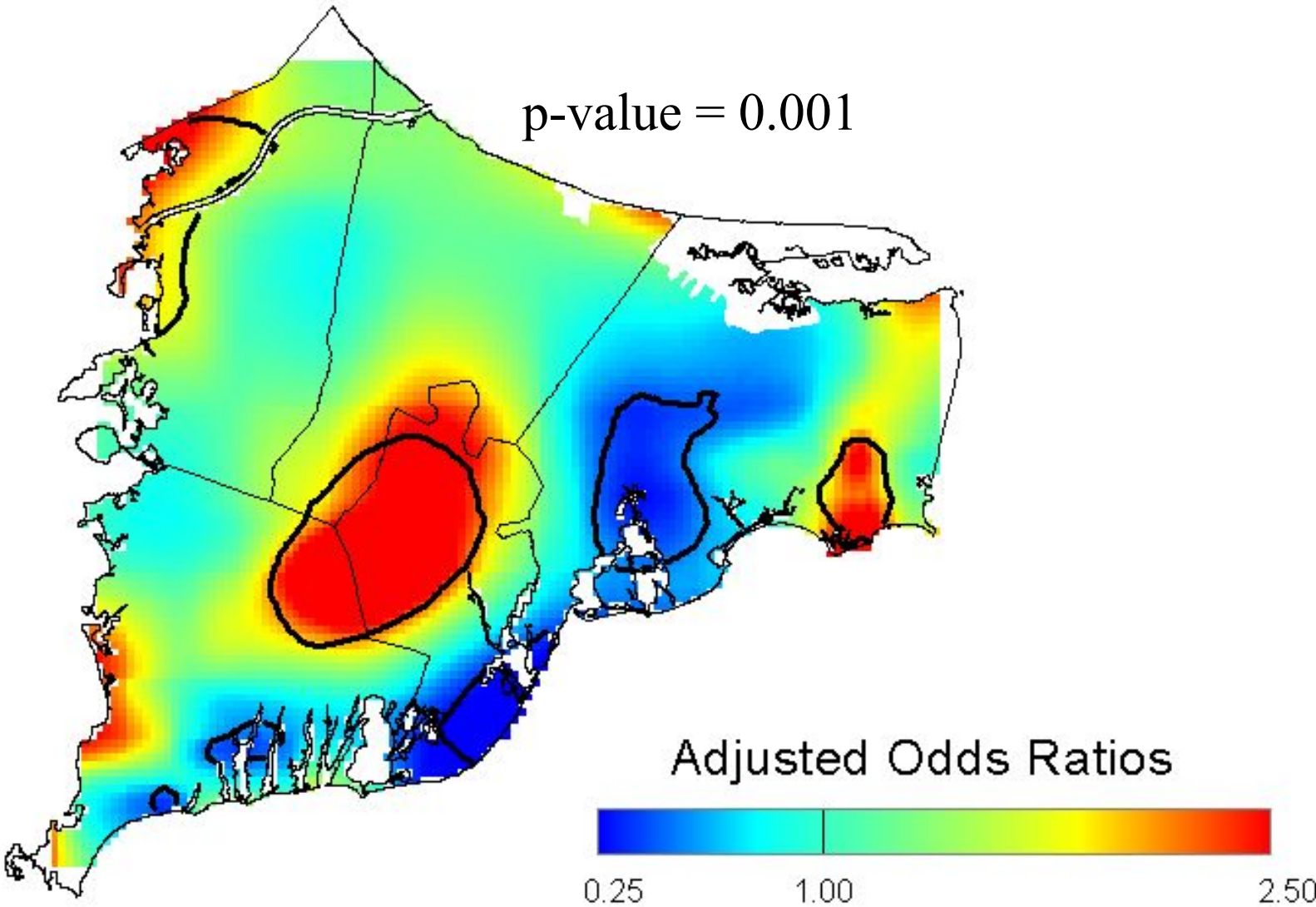
0.25 1.00 2.50



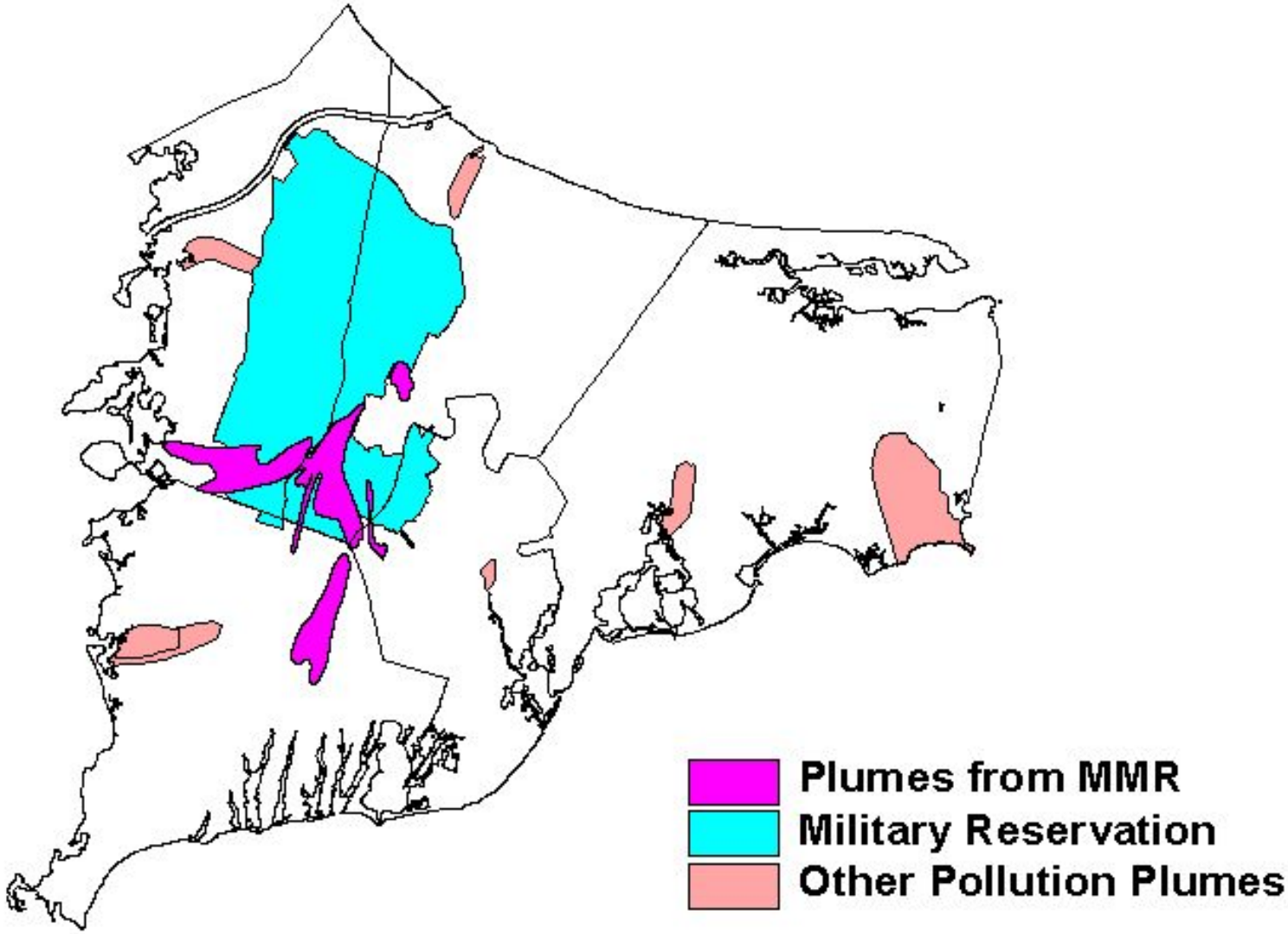
Spatial Confounding



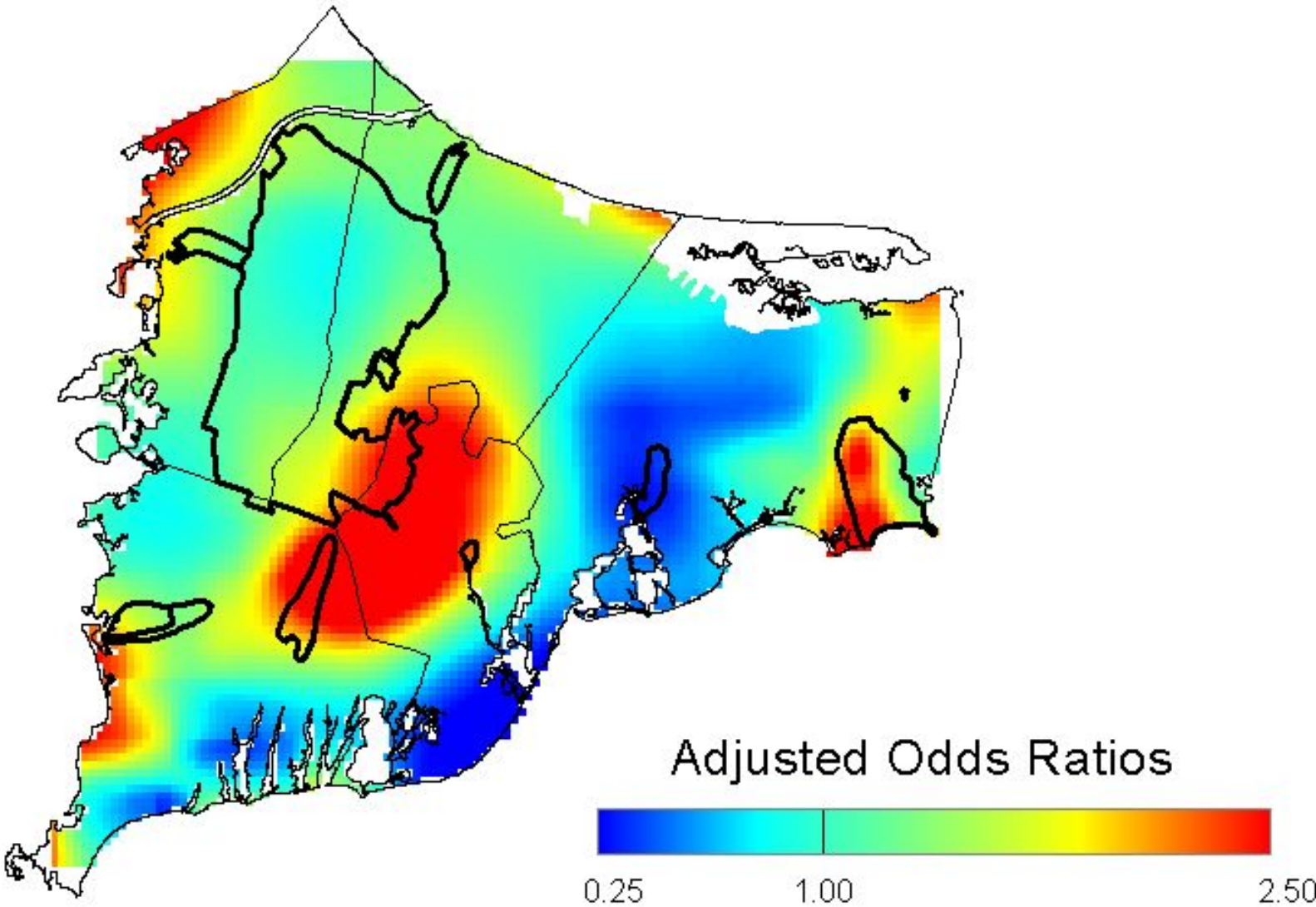
Significant Risk Areas



MMR and Plumes



Proximity to MMR and Plumes



Discussion

- Can GIS help us better understand bioavailability?
- GPS can be used to develop time-location information.
- How can the GAM mapping methods be applied?
- The model may be able to predict dose based on location and other personal information.
- Environmental, biological data is needed.

Acknowledgements

Thomas Webster, Janice Weinberg, Ann Aschengrau, and David Ozonoff.

This work was supported by Superfund Basic Research Grant 1P42ES 07381.